Columbia DBMI's OHDSI "Summer School in Observational Health Data Science & Informatics, AI, and Real-World Evidence"

George Hripcsak, MD, MS

Director, OHDSI Coordinating Center

Biomedical Informatics, Columbia University

From Health Data to Discovery: Learn Real-World Evidence at Columbia's OHDSI Summer School

Leaders within the OHDSI community will host a **Summer School in Observational Health Data Science & Informatics, AI, and Real World Evidence** this July 14-18 within the Department of Biomedical Informatics at Columbia University.

George Hripcsak, Patrick Ryan, Anna Ostropolets and Karthik Natarajan will serve as faculty for this session, which will be limited to 30 participants to ensure a high level of interaction and personal support.

Nearly 500 people joined the 2024 OHDSI Global Symposium last October. Columbia University serves as the coordinating center for OHDSI.

About the Summer School

The Columbia OHDSI Summer School provides health professionals, researchers, and industry practitioners with an immersive, hands-on training to working with real-world health data and generating real-world evidence (RWE). Participants will explore the types of healthcare data captured during routine clinical care—such as electronic health records and administrative claims—and learn how to standardize these data using the OMOP Common Data Model to support collaborative, distributed research as part of a data network.

Over the course of the week, participants will engage with three real-world analytic use cases:

- **Clinical characterization** – using descriptive epidemiology to study disease natural history and treatment patterns
- **Population-level estimation** – applying causal inference to assess drug safety and comparative effectiveness
- **Patient-level prediction** – leveraging machine learning for early disease detection and precision medicine

Participants will be guided through the full RWE study lifecycle: from designing observational studies tailored to each use case, to applying ope n-source tools form the OHDSI community, and executing analyses across real-world data sources.

The curriculum combines foundational lectures on analytical methods with hands-on, interactive, faculty-led group exercises. In addition, participants will have dedicated time to develop and advance their own study concepts with personalized feedback and mentoring.

Goal

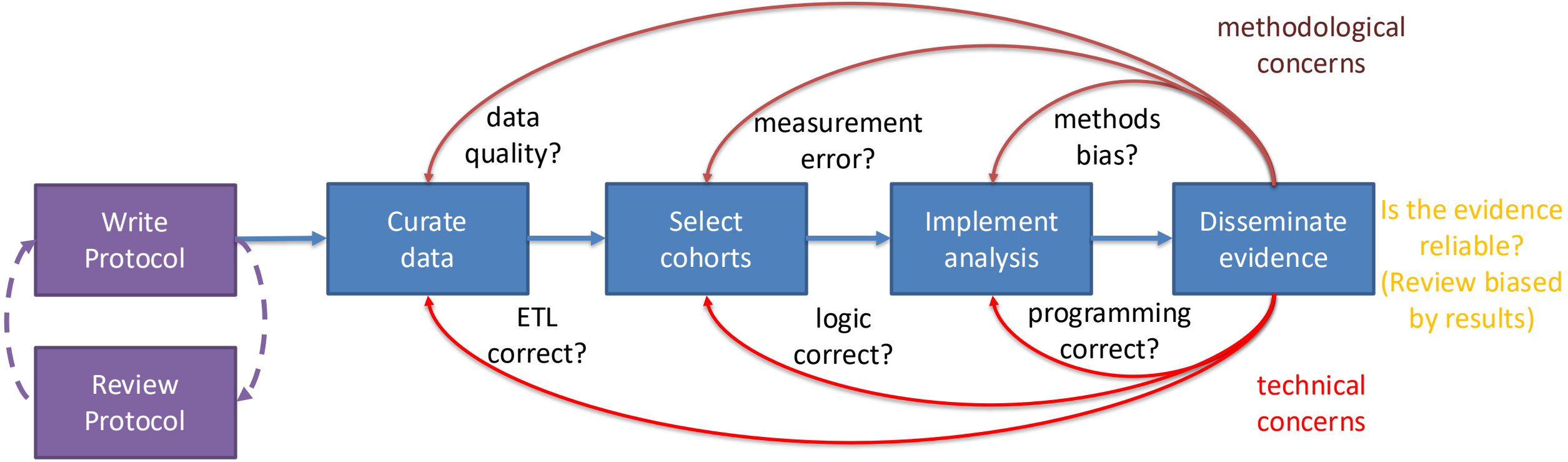
- Gain foundational and practical knowledge in observational research
- Hands-on experience conceiving, designing, implementing, running, and reviewing the results of a OHDSI network study

Audience

- Interest in observational research
- Willing to work hard for a week
- Varying levels of expertise
  - First timers
  - Statisticians and epidemiologists
  - Clinicians
  - Software engineers
  - Informaticians

Walking through the steps of reliable research

Does the study provide an unbiased effect estimate?
Are the findings generalizable to the population of interest?



Can the study be fully reproduced?
Does the analysis actually do what the protocol said it would do?

Activities

- Hands-on experience with multiple large databases (7-30M)
  - Clinical characterization
  - Population-level estimation
  - Patient-level prediction
- From designing observational studies tailored to each use case, to applying open-source tools from the OHDSI community, and executing analyses across real-world data sources
- Foundational lectures on analytical methods with hands-on, interactive, faculty-led group exercises

Activities

- A tutorial that evolves into a study-a-thon
- Students generated their own hypotheses
  - Selected six for group execution

Resources

- Faculty experienced in observational research and OHDSI tools
- Columbia DBMI space
- Exploit the AWS platform
  - Database server
  - Individual student virtual machines
  - Crank up the power (many cores) for large-scale calculations
- EHR and claims databases with 7 to 30 million persons

This year’s experience

- Was held July 2025
- 30 participants
  - International (12 hr time zone), across the US, local
  - Clinicians, epidemiologists, informaticians, analysts, students, faculty
- Fully engaged
  - Remained through the course other than one or two who knew they had to leave early


EXIT


mer School Agenda

Faculty

George Hripcsak, MD, MS
Vivian Beaumont Allen Professor,
Columbia Biomedical Informatics

Patrick Ryan, PhD
Adjunct Assistant Professor,
Columbia Biomedical Informatics
Johnson & Johnson

Anna Ostropolets, PhD
Adjunct Assistant Professor,
Columbia Biomedical Informatics
Johnson & Johnson

Karthik Natarajan, PhD
Assistant Professor,
Columbia Biomedical Informatics

Columbia DBMI Summer School Agenda

Goal and data | Study design | Do phenotype | Do analysis | Results

	Monday	Tuesday	Wednesday	Thursday	Friday
8am-9am	**IT test:** VM and ATLAS				
900am-1200pm	**Lecture:** Introduction to observational network research **Lecture:** Understanding administrative claims **Lecture:** Understanding electronic health records **Exercise:** Mapping a Patient journey	**Lecture:** Designing an observational network study: Characterization **Lecture:** Designing an observational network study: Estimation	**Lecture:** Phenotype development **Demo:** Build cohorts in ATLAS	**Lecture:** Implementing an observational network study using OHDSI framework **Exercise:** Review CohortDiagnostics results for your own cohorts **Demo:** Create input specifications, execute package	**Lecture:** Interpreting results from an observational network study and building trust in evidence - Openness and verification - Diagnostics - Evidence Synthesis **Demo:** Review Rshiny app
12pm-1pm	Lunch	Lunch	Lunch	Lunch	Lunch
100pm-500pm	**Lecture:** OMOP Common Data Model **Exercise:** Exploring the OHDSI Standardized Vocabularies in ATLAS **Exercise:** Exploring patient profiles in ATLAS **430pm: IT test**: execute R package	**Lecture:** Designing an observational network study: Prediction **Exercise:** Framing your research question using OHDSI standard questions **Exercise:** Evaluating data network fitness-for-use using ATLAS/Data Sources	**Exercise:** Build cohorts in ATLAS **Lecture:** Phenotype evaluation - CohortDiagnostics - PheValuator - KEEPER **Exercise:** Review CohortDiagnostics results	**Exercise:** Create input specifications, execute Strategus	**Exercise:** Review RShiny app and present findings
5pm-6pm	Group photo from DBMI Terrace	Break	Evening on your own	Evening on your own	School's out for summer!
6pm-9pm					

DBMI Summer School Agenda - Monday

Time	Topic	Lead
8:00am-9:00am	**IT test and troubleshooting**: log into VM and ATLAS	Karthik
9:00am-9:45am	**Lecture:** Introduction to observational network research (introduce running example: ACE v. ARB for AMI and angioedema)	George
9:45am-10:30am	**Lecture:** Understanding administrative claims (Merative)	Patrick
10:30am-10:45am	Break	
10:45am-11:30am	**Lecture:** Understanding electronic health records (CUIMC)	Karthik
11:30am-12:00pm	**Exercise:** Mapping a Patient journey	Patrick
12:00pm-1:00pm	Lunch	
1:00pm-2:00pm	**Lecture:** OMOP Common Data Model	Patrick
2:00pm-2:45pm	**Exercise:** Exploring the OHDSI Standardized Vocabularies in ATLAS	Anna
2:45pm-3:00pm	Break	
3:00pm-3:45pm	**Lecture:** Extract, transform, load (ETL) to OMOP CDM and Evaluating Data Quality	Karthik
3:45pm-4:30pm	**Exercise:** Exploring patient profiles in ATLAS	Patrick
4:30pm-5:00pm	**430pm: IT test #2**: execute R package	Karthik
5:00pm-5:15pm	Group photo from DBMI Terrace	Craig
5:15pm-midnight	Evening on your own	

DBMI Summer School Agenda - Tuesday

Time	Topic	Lead
9:00am-10:00am	**Lecture:** Designing an observational network study: Characterization	Anna
10:00am-10:30am	**Exercise:** Framing your research question using OHDSI standard questions	Patrick
10:30am-10:45am	Break	
10:45am-12:00pm	**Lecture:** Designing an observational network study: Estimation using Comparative Cohort design	George
12:00pm-1:00pm	Lunch	
1:00pm-1:30pm	**Lecture:** Designing an observational network study: Estimation using SelfControlledCaseSeries	George
1:30pm-2:00pm	**Exercise:** Framing your research question using OHDSI standard questions	Anna
2:00pm-3:00pm	**Lecture:** Designing an observational network study: Patient-level Prediction	Patrick
3:00pm-3:30pm	**Exercise:** Framing your research question using OHDSI standard questions	Patrick
3:30pm-3:45pm	Break	
3:45pm-4:45pm	**Exercise:** Evaluating data network fitness-for-use using ATLAS/Data Sources	Karthik
4:45pm-5:00pm	**Exercise:** Network study team formation	Patrick
5:00pm-6:00pm	Break	
6:00pm-10:00pm	Summer school social night excursion	

DBMI Summer School Agenda - Wednesday

Time	Topic	Lead
9:00am-9:45am	**Lecture:** Overview on phenotype development and evaluation	Patrick
9:45am-10:30am	**Demo:** Introduction to ACE conceptsets	Anna
10:30am-10:45am	Break	
10:45am-12:00pm	**Exercise:** Build your own conceptsets in ATLAS	Anna
12:00pm-1:00pm	Lunch	
1:00pm-2:00pm	**Demo:** Build ACE cohorts in ATLAS	Patrick
2:00pm-2:45pm	**Exercise:** Build your own cohorts in ATLAS	Patrick
2:45pm-3:00pm	Break	
3:00pm-4:00pm	**Lecture:** Phenotype evaluation - CohortDiagnostics - PheValuator - KEEPER	George
4:00pm-5:00pm	**Exercise:** Review ACE CohortDiagnostics results and your own cohort counts	Anna
5:00pm-midnight	Summer School Social Night local pizza	

DBMI Summer School Agenda - Thursday

Time	Topic	Lead
9:00am-9:30am	**Lecture:** Implementing an observational network study using OHDSI framework	George
9:30am-10:30am	**Exercise:** Review CohortDiagnostics results for your own cohorts	Patrick
10:30am-10:45am	Break	
10:45am-12:00pm	**Demo:** Create ACE input specifications and execute Strategus package	Anna
12:00pm-1:00pm	Lunch	
1:00pm-2:45pm	**Exercise:** Create your own input specifications	George
2:45pm-3:00pm	Break	
3:00pm-4:00pm	**Exercise:** Execute your own package	Karthik
4:00pm-5:00pm	**Exercise:** 10-minute lightning talk: present your protocol	Patrick
5:00pm-midnight	Evening on your own	

DBMI Summer School Agenda - Friday

Time	Topic	Lead
9:00am-9:30am	**Lecture:** Interpreting results from an observational network study and building trust in evidence - Openness and verification - Diagnostics	George
9:30am-10:30am	**Demo:** Review ACE results in RShiny	Patrick
10:30am-10:45am	Break	
10:45am-12:00pm	**Exercise:** Interpret your own study results in RShiny	Patrick
12:00pm-1:00pm	Lunch	
1:00pm-2:45pm	**Exercise:** Interpret your own study results in Rshiny (cont)	Patrick
2:45pm-3:00pm	Break	
3:00pm-5:00pm	**Exercise:** 10-minute lightning talk: present your study findings	Anna/George/ Karthik/Patrick
5:00pm onward	School's out for summer!	

Results

- The group produced six network studies
  - Two characterization
  - Two population-level estimation
  - Two prediction
  - Three are going on to further expansion
- Team building
- Some group excursions for those from outside New York


OHDSI


SANDWICHES
B'S COMBO $19.85
LAUNDERETTE $19.88
REFRIGERATOR $19.91
POOL HALL
MARRIOTT MARQUIS
OPENING
Levi's
OHDSI


BE KIND
DANGER
OTHER HALF


LITTLE, BIG, AND FAR
ABRAHAM'S BOYS
DOC NYC SELECTS
20 YEARS OF IFC CENTE
EXIT ONLY
ENTRY ONLY
ATM

Reviews

- Main question: How satisfied are you with the overall summer school experience?
  - 100% top rating
- Met my expectations, would recommend, relevant, appropriate, balance of theory and practice, made sense
  - Mostly strongly with a few somewhat agree, nothing lower
- What I can now accomplish…
  - Mostly strongly with a few somewhat agree, nothing lower
- Instructors
  - Almost unanimous strongly agree

Reviews

- Most valuable: Understanding how a study based on the OHDSI is set up from task definition, cohort definitions and diagnostics, running experiments and interpreting results, guided by the technical tools provided by OHDSI
- Improvements: More time on CDM, more technical, less technical, I don’t like sandwiches

Testimonials

- "**The OHDSI Summer School exceeded all expectations. The course offered a deep, hands-on dive** into real-world data methodology, led by world-class instructors and supported by an incredibly open, collaborative community. Working hands-on with a team to explore a real research question **brought the full OHDSI workflow to life**. **I highly recommend this course** to anyone working with health data, especially clinical scientists and data scientists eager to strengthen their skills in transparent, high-quality RWD research."
  - David Bard, U of Oklahoma
- "**I particularly appreciated the OHDSI system in conducting the observational study to generate real-world evidence,** as well as the methodology that handled hundreds of covariates, and the application of a negative control outcome to validate the research results. Through this summer school course, I gained a deeper understanding of OHDSI and its potential to generate real-world evidence."
  - Long-Shen Chen, National Taipei University of Technology

Next year

- Scheduled for 2026 June 22-26
- Registration
  - https://columbiauniversity1.regfox.com/columbia-ohdsi-summer-school-2026
- Limited to 30